Package: riv (via r-universe)

August 21, 2024

Title Robust Instrumental Variables Estimator

Version 2.0-5

Date 2018-05-023

Author Gabriela Cohen-Freue and Davor Cubranic, with contributions
from B. Kaufmann and R.H. Zamar

Maintainer Gabriela Cohen-Freue <gcohen@stat.ubc.ca>

Depends MASS, rrcov, quantreg

Description Finds a robust instrumental variables estimator using a
high breakdown point S-estimator of multivariate location and
scatter matrix.

License GPL-2

NeedsCompilation no

RoxygenNote 6.0.1

Date/Publication 2018-05-24 03:09:12 UTC

Repository https://gcohenfr.r-universe.dev

RemoteUrl https://github.com/cran/riv

RemoteRef HEAD

RemoteSha 26c4574b86d8e373dac1958e73725873bbe6d492

Contents

1

---

riv-package   *Robust instrumental variables estimator package.*

Description

This package contains tools to find a robust instrumental variables estimator based on a high breakdown point S-estimator of location and covariance.

Details

riv(Y, Xend, Xex = NULL, Zinst, dummies = NULL, method = c('S-est', 'SD-est', 'MCD-est', 'classical'), nsamp = 500, bdp = 0.5)

finds a robust instrumental variables estimator using a high breakdown point S-estimator of location and covariance.

Author(s)

G.V. Cohen-Freue <gcohen@stat.ubc.ca>

D. Cubranic <cubranic@stat.ubc.ca>

with contributions from B. Kaufmann <beat.kaufmann@epfl.ch> and R.H. Zamar <ruben@stat.ubc.ca>

References

LOPUHAA,H.P. (1989). *On the Relation between S-estimators and M-estimators of Multivariate Location and Covariance. Ann. Statist.* **17** 1662-1683.

COHEN-FREUE, G.V., ORTIZ-MOLINA, H., and ZAMAR, R.H. (2012) *A Natural Robustification of the Ordinary Instrumental Variables Estimator.* Submitted to *Biometrics*.

Examples

```
## load data earthquake: the first column is the response Y, the second
## the endogenous variable X and the third column is the instrument
## Zinst.
data(earthquake)
riv(earthquake[,1], earthquake[,2], NULL, earthquake[,3])
```

---

earthquake *Alaskan earthquakes data*

Description

The dataset contains information about 62 Alaskan earthquakes that occured between 1969-1978 (Fuller, 1987). The goal is to see how the earthquake strength, measured in terms of the true value of the body waves, x^*, impacts on the amplitude of the surface waves of the earthquake (Y). However, we do not observe x^* but X, which is the logarithm of the seismogram amplitude of longitudinal body waves measured at some observation stations, i.e., $x_i = x_i^* + u_i$, for $i = 1, \ldots, 62$, where u_i is the measurement error of each observation. Thus, in the regression $Y_i = a + b * X_i + e_i$, the covariate X is an endogenous covariate. We can consistently estimate the regression parameters using instrumental variables estimators. The logarithm of maximum seismogram trace amplitude at short distance, W, can be used as an instrument.

The first column in the dataset is the response (Y), the second column is the endogenous variable (X), and the third column is the instrument (W).

Usage

```
data(earthquake)
```

Format

A data frame with 62 observations on the following 3 variables.

Y a numeric vector of the logarithm of the seismogram amplitude of 20 second waves.

X a numeric vector of the logarithm of the seismogram amplitude of longitudinal body waves.

W a numeric vector of the logarithm of maximum seismogram trace amplitude at short distance.

Source

FULLER,W.A. (1987). *Measurement Error Models*. Wiley, New York.

References

COHEN-FREUE,G.V. and ZAMAR,R.H. (2005). *A Robust Instrumental Variables Estimator*.

Examples

```
data(earthquake)
plot(earthquake$X, earthquake$Y, xlab="X", ylab="Y")
```

---

mortality	*Mortality rate data*

Description

The dataset contains information about the mortality rate from 60 U.S. cities using aggregate information from the year 1969/70.

Usage

```
data(mortality)
```

Format

A data frame with 60 observations on the following 8 variables.

MO70 a numeric vector of the total mortality (number of deaths per 1000 people) from 1970.

MAGE a numeric vector of the median age of the population (in years) from 1969.

CI68 a numeric vector of the number of packs of cigarettes per year per person.

MDOC a numeric vector of the density of medical doctors (number of medical doctors per 100,000 people).

DENS a numeric vector of the percentage of households with more than 1.5 persons per room.

NONW a numeric vector of the fraction of the non-white population.

EDUC a numeric vector of the percentage of the population over age 25 having a high-school diploma.

IN69 a numeric vector of the median income from 1969.

Source

CROCKER,D.T. et al. (1979). *Methods Development for Assessing Air Pollution Control Benefits, Vol. 1. Experiments in the Economics of Epidemiology.* EPA-600/5-79-001a. Springfield, VA; National Technical Information Service.

Examples

```
data(mortality)
```

---

riv   *Robust instrumental variables estimator*

Description

Finds robust instrumental variables estimator using high breakdown point multivariate location and scatter matrix S-estimators.

Usage

```
riv(Y, Xend, Xex=NULL, Zinst, dummies=NULL,
    method = c('S-est', 'SD-est', 'MCD-est', 'classical'))
```

Arguments

Y	vector of responses.
Xend	matrix of the endogenous variables, i.e. covariates that are correlated with the regression's error term.
Xex	matrix of the exogenous variables, i.e. covariates that are uncorrelated with the regression's error term. Default = NULL, i.e. no exogenous variables are present in the model.
Zinst	matrix of instruments, variables correlated with the endogenous covariates, but uncorrelated with the error term. The number of instrumental variables needs to be larger than or equal to the number of endogenous covariates.
dummies	matrix of exogenous dummy covariates, i.e., where each D_i are 0–1 vectors.
method	the method to be used. The "S-est" method (default) is based on the S-estimator of multivariate location and covariance, and "classical" method is based on the sample mean and covariance and the resulting estimator is equivalent to the two-stage least squares estimator (2SLS). See Details section.

Details

For method "S-est", RIV is constructed using the robust multivariate location and scatter S-estimator based on the Tukey's biweight function (see CovSest).

If RIV is computed using the S-estimator, its variance-covariance matrix is estimated based on the empirical influence function. See references for more details.

For method "SD-est", RIV is constructed using the Stahel-Donoho's robust multivariate location and scatter estimator (see CovSde).

For method "MCD-est", RIV is constructed using the Minimum Covariance Determinant (MCD) robust multivariate location and scatter estimator (see CovMcd).

For method "classical", the estimator is the classical instrumental variables estimator based on the sample mean and sample variance-covariance matrix (also known as the two-stage least squares estimator, 2SLS).

If the model contains dummy variables (i.e., dummies != NULL), RIV is computed using an iterative algorithm called "L_1-RIV". Briefly, L_1-RIV estimates the coefficients of the dummies using an L_1-estimator and the coefficients of the continuous covariates using the original RIV. See Cohen Freue et al. for more details.

Value

A list with components:

Summary.Table	Matrix of information available about the estimator. It contains regression coefficients, and, for method = "S-est" and "classical" only, columns for the standard error, t-statistic, and p-value.
VC	estimated variance-covariance matrix, computed only if method = "S-est" or "classical"
MD	Squared Mahalanobis distances of each observation to the multivariate location S-estimator with respect to the scatter S-estimator (only computed if method = "S-est").
MSE	vector of three components, computed only if method = "S-est" or "classical":

- sigma.hat1: the mean square error estimation;
- sigma.hat2: the mean square error estimation taking into account the weights associated to each observation (only computed if method = "S-est" and coefficents of endogenous variables are exactly identified, i.e., the number of instruments is equal to the number of endogenous variables);
- sigma.hat3: the squared median absolute deviation (only computed if method = "S-est").

weight	the weights assigned by RIV to each observation (only computed if method = "S-est" and the number of instruments is equal to the number of endogenous variables.

References

LOPUHAA H.P. (1989). *On the Relation between S-estimators and M-estimators of Multivariate Location and Covariance. Ann. Statist.* **17** 1662-1683.

COHEN-FREUE, G.V., ORTIZ-MOLINA, H., and ZAMAR, R.H. (2012) *A Natural Robustification of the Ordinary Instrumental Variables Estimator.* Submitted to *Biometrics*.

See Also

[CovSest](), [CovSde](), [CovMcd]()

Examples

```
## load data earthquake: the first column contains the response (Y), the
## second the endogenous variable (X) and the third column is the
## instrument (W).
data(earthquake)
riv.eq <- riv(earthquake$Y,earthquake$X,NULL,earthquake$W)
```

```
## plot of the RIV estimates and the outlying observations are
## identified by filled points
plot(earthquake$X,earthquake$Y,xlab="X",ylab="Y",cex=1.5)
abline(riv.eq$Summary.Table[,1])
outliers <- which(sqrt(riv.eq$MD)>sqrt(qchisq(0.99, 3)))
text(earthquake[outliers,2],
     earthquake[outliers,1],
     outliers,
     pos=c(4,4,4,2))

points(earthquake[outliers,2],
       earthquake[outliers,1],
       cex=1.5,pch=19)

## Weights given by RIV to each observation as a function of the square
## root of the Mahalanobis distances (d) of each observation to the
## multivariate location and covariance S-estimator (computed with
## CovSest in rrcov)
plot(sqrt(riv.eq$MD),riv.eq$weight,xlab="d",ylab="RIV's Weights",cex = 1.5)
abline(h=sqrt(qchisq(0.99, 3)))
text(sqrt(riv.eq$MD)[outliers],
     riv.eq$weight[outliers],
     outliers,
     pos=c(2, 1, 1, 4))
points(sqrt(riv.eq$MD)[outliers],
       riv.eq$weight[outliers],
       cex=1.5, pch=19)


## load data mortality
data(mortality)
Y <- as.matrix(mortality[,1])          ## M070
Xex <- as.matrix(mortality[,c(2,3,5,6)]) ## MAGE,CI68,DENS,NONW
Xend <- as.matrix(mortality[,4])        ## MDOC
colnames(Xend) <- colnames(mortality)[4]
Zinst <- as.matrix(mortality[,7:8])     ## EDUC,IN69

## Classical instrumental variables estimator
riv(Y, Xend, Xex, Zinst, method="classical")
```

Index